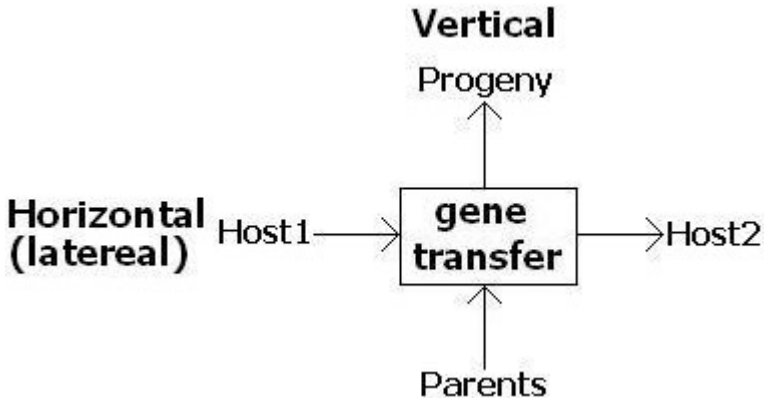
Horizontal genetic flow by sperm cells: evidences and hypothesis

Andrew Kuznetsov

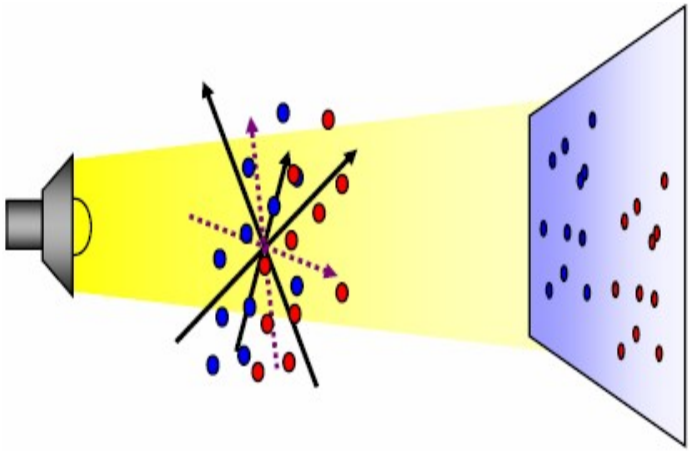
Freiburg, Germany

Content

- Communication, modularity and complexity: from ancient to modern world
  - ancient HGT
  - examples
  - HGT and biological complexity
  - role of RNA
- Dual role of sperm cells
  - the RNA (DNA) uptake by sperm cells
  - the reverse transcriptase activity in SP
  - propagation of ‘transgenes’
- Repression of transposons by Argonaute family proteins: all or nothing?
  - in *Drosophila* and mouse
  - Ping-Pong mechanism
  - stochastic activity of transposones
  - paramutations
- Oracle language of RNA
  - RNA logic evaluator
  - RNA-guided recombination in ciliates

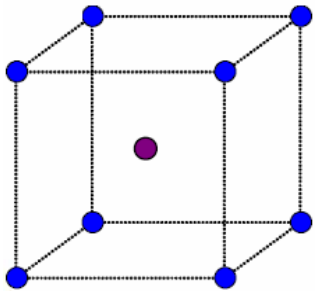


Data acquisition and their interpretation



Principal Component Analysis (PCA), Multivariate Projection - SIMCA-P + 12

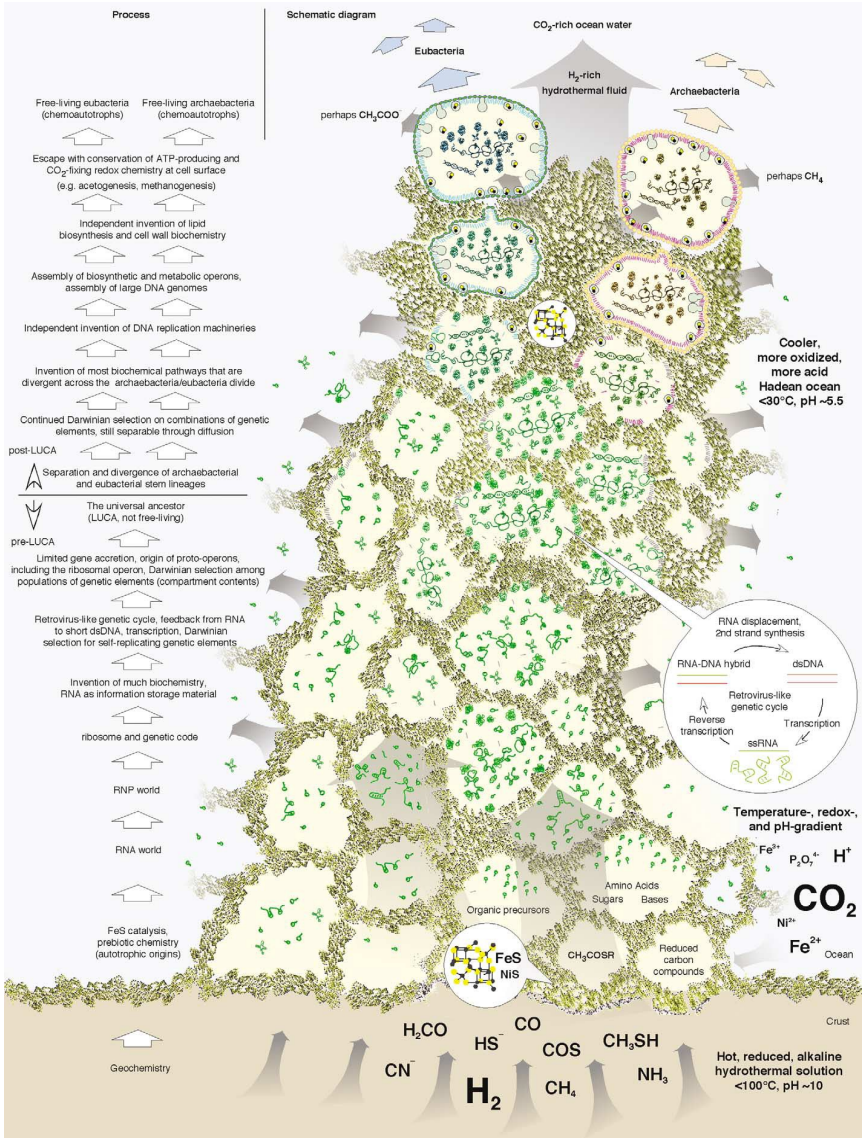
Ing-Marie Olsson, http://www.umetrics.com



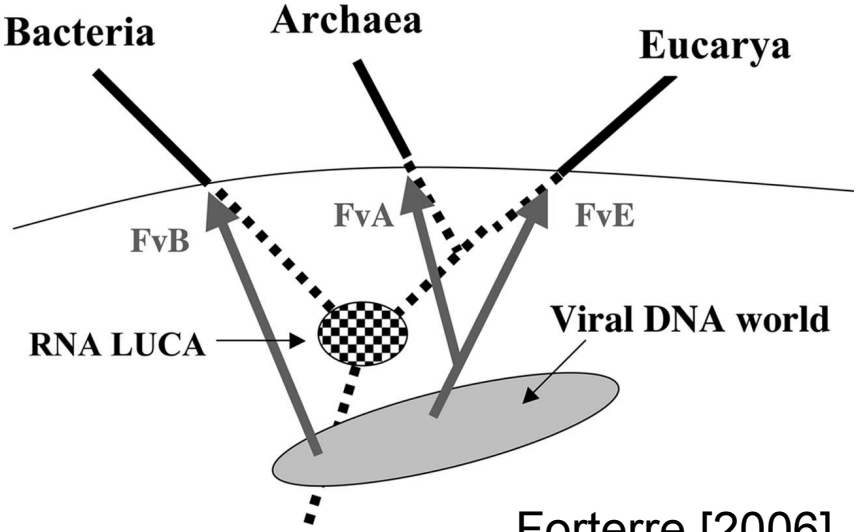
Design of experiments (DOE) - MODDE 8

Proof of concept by artificial design:
'*What I cannot create I do not understand*' - Erwin Schrödinger

3.5-3.8 billion years ago.. The Russell-Martin-Koonin scenario of origin of life



Wachtershauser [1988], Russell and Martin [2004], Koonin and Martin [2005]



Forterre [2006]

'*Arguments that attempt to extrapolate from modern biochemistry back to the origin of life are futile*' - Steven Benner

Short literature overview: HGT

ANCIENT HORIZONTAL GENE TRANSFER

James R. Brown

The cornerstone of Charles Darwin's theory of evolution is the vertical inheritance of traits from parent to offspring across successive generations. However, molecular evolutionary biologists have shown that extensive horizontal (also known as lateral) gene transfer (HGT) can occur between distantly related species. Comparative sequence analyses of genomes indicates that the universal tree of life might be at risk because of pervasive, ancient HGT. Considerable debate now ensues about the role of HGT in genome evolution. At stake are a fundamental understanding of how life evolved and a deeper knowledge of the functioning of all genomes, including that of humans.

Unsupervised Learning in Detection of Gene Transfer

Lutz Hamel[1], Neha Nahar[1], Maria S. Poptsova[2], Olga Zhaxybayeva[3], J. Peter Gogarten[2]

[1] Department of Computer Science and Statistics, University of Rhode Island, USA
[2] Department of Molecular and Cell Biology, University of Connecticut, USA
[3] Department of Biochemistry and Molecular Biology, Dalhousie University, Canada

hamel@cs.uri.edu, nnahar@cs.uri.edu, maria.poptsova@uconn.edu, olgazh@dal.ca, gogarten@uconn.edu

Abstract

The tree representation as a model for organismal evolution has been in use since before Darwin. However, with the recent unprecedented access to biomolecular data it has been discovered that, especially in the microbial world, individual genes making up the genome of an organism give rise to different and sometimes conflicting evolutionary tree topologies. This discovery calls into question the notion of a single evolutionary tree for an organism and gives rise to the notion of an evolutionary consensus tree based on the evolutionary patterns of the majority of genes in a genome embedded in a network of gene histories. Here we discuss an approach to the analysis of genomic data of multiple genomes using bipartition spectral analysis and unsupervised learning. An interesting observation is that genes within genomes that have evolutionary tree topologies that are in substantial conflict with the evolutionary consensus tree of an organism point to possible horizontal gene transfer events which often delineate significant evolutionary events.

PNAS

Highways of gene sharing in prokaryotes

Robert G. Beiko, Timothy J. Harlow, and Mark A. Ragan*

Institute for Molecular Bioscience and Australian Research Council Centre in Bioinformatics, The University of Queensland, Brisbane 4072, Australia

Edited by Carl R. Woese, University of Illinois at Urbana–Champaign, Urbana, IL, and approved August 17, 2005 (received for review May 16, 2005)

The extent to which lateral genetic transfer has shaped microbial genomes has major implications for the emergence of community structures. We have performed a rigorous phylogenetic analysis of >220,000 proteins from genomes of 144 prokaryotes to determine the contribution of gene sharing to current prokaryotic diversity, and to identify "highways" of sharing between lineages. The inferred relationships suggest a pattern of inheritance that is largely vertical, but with notable exceptions among closely related taxa, and among distantly related organisms that live in similar environments.

lateral genetic transfer | microbial genomes | molecular phylogeny

methods on a subset of available taxa with restricted phylogenetic (14) or environmental (15) distributions. However, the set of sequenced genomes represents taxa that are phylogenetically and ecologically diverse, and can be used to test broad hypotheses about the sharing of genes. Here we apply rigorous phylogenetic methods to annotated proteins from 144 completely sequenced genomes sampled across 15 phyla of prokaryotes. We derive a reference supertree from 22,432 orthologous protein families, and by comparing individual protein trees with this reference tree, we infer for this data set the frequency and phyletic extent of LGT, the taxa implicated as partners in LGT events, and the tendency of different cellular functions to be subjected to transfer.

PNAS

Decoding the genomic tree of life

Anne B. Simonson*†, Jacqueline A. Servin‡, Ryan G. Skophammer*, Craig W. Herbold‡, Maria C. Rivera*†, and James A. Lake*†‡§¶

‖Molecular Biology Institute, Departments of *Molecular, Cell, and Developmental Biology and §Human Genetics, and †National Aeronautics and Space Administration Astrobiology Institute, University of California, 242 Boyer Hall, Los Angeles, CA 90095

Genomes hold within them the record of the evolution of life on Earth. But genome fusions and horizontal gene transfer (HGT) seem to have obscured sufficiently the gene sequence record such that it is difficult to reconstruct the phylogenetic tree of life. HGT among prokaryotes is not random, however. Some genes (informational genes) are more difficult to transfer than others (operational genes). Furthermore, environmental, metabolic, and genetic differences among organisms restrict HGT, so that prokaryotes preferentially share genes with other prokaryotes having properties in common, including genome size, genome G+C composition, carbon utilization, oxygen utilization/sensitivity, and temperature optima, further complicating attempts to reconstruct the tree of life. A new method of phylogenetic reconstruction based on gene presence and absence, called conditioned reconstruction, has improved our prospects for reconstructing prokaryotic evolution. It is also able to detect past genome fusions, such as the fusion that appears to have created the first eukaryote. This genome fusion between a deep branching eubacterium, possibly an ancestor of the cyanobacterium and a proteobacterium, with an archaeal eocyte (crenarchaea), appears to be the result of an early symbiosis. Given new tools and new genes from relevant organisms, it should soon be possible to test current and future fusion theories for the origin of eukaryotes and to discover the general outlines of the prokaryotic tree of life.

plants and animals, namely the enigmatic evolution of prokaryotes and the emergence of eukaryotes.

The origin of the eukaryotes was a milestone in the evolution of life, because eukaryotes are utterly different from prokaryotes in their spatial organization. Eukaryotes, for example, possess an extensive system of internal membranes that traverse the cytoplasm and enclose organelles, including the mitochondrion, chloroplast, and nucleus. This compartmentalization has required a number of unique eukaryotic innovations. The most dramatic innovation is the nucleus, a specific compartment for storing and transcribing DNA, for processing DNA and RNA, and possibly even for translating mRNAs (18). The nucleus is unique to eukaryotes, hence it and the nuclear genome are the defining characters for which eukaryotes are named (eu, good or true; karyote, kernel, as in nucleus).

The prokaryotes, with their simple cellular organization, are generally thought to have preceded the eukaryotes (although see ref. 19). Which prokaryotic groups branched first, however, is not clear, because the root of the tree of life is uncertain and in flux due to a concern that artifacts of phylogenetic reconstruction may have unduly influenced the location of even the root that has the most experimental support (20, 21).

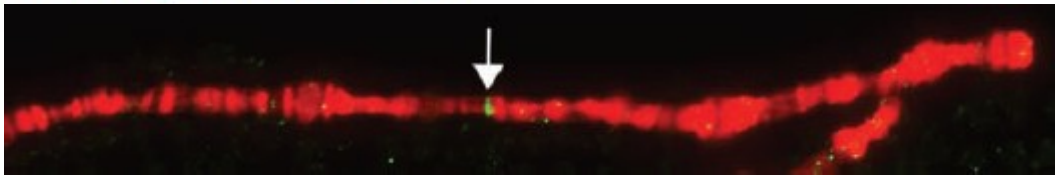
The HGT Revolution

The possibility of analyzing complete genomes awakened inter-

Widespread Lateral Gene Transfer from Intracellular Bacteria to Multicellular Eukaryotes

Julie C. Dunning Hotopp,[1*†‡] Michael E. Clark,[2*] Deodoro C. S. G. Oliveira,[2] Jeremy M. Foster,[3] Peter Fischer,[4] Mónica C. Muñoz Torres,[5] Jonathan D. Giebel,[2] Nikhil Kumar,[1‡] Nadeeza Ishmael,[1‡] Shiliang Wang,[1] Jessica Ingram,[3] Rahul V. Nene,[1§] Jessica Shepard,[1‖] Jeffrey Tomkins,[5] Stephen Richards,[6] David J. Spiro,[1] Elodie Ghedin,[1,7] Barton E. Slatko,[3] Hervé Tettelin,[1‡¶] John H. Werren[2¶]

Although common among bacteria, lateral gene transfer—the movement of genes between distantly related organisms—is thought to occur only rarely between bacteria and multicellular eukaryotes. However, the presence of endosymbionts, such as *Wolbachia pipientis*, within some eukaryotic germlines may facilitate bacterial gene transfers to eukaryotic host genomes. We therefore examined host genomes for evidence of gene transfer events from *Wolbachia* bacteria to their hosts. We found and confirmed transfers into the genomes of four insect and four nematode species that range from nearly the entire *Wolbachia* genome (>1 megabase) to short (<500 base pairs) insertions. Potential *Wolbachia*-to-host transfers were also detected computationally in three additional sequenced insect genomes. We also show that some of these inserted *Wolbachia* genes are transcribed within eukaryotic cells lacking endosymbionts. Therefore, heritable lateral gene transfer occurs into eukaryotic hosts from their prokaryote symbionts, potentially providing a mechanism for acquisition of new genes and functions.



Evidence supporting Wolbachia/host HGT by fluorescence microscopy. DNA in the polytene chromosomes of *D.ananassae* were stained with propidium iodide (red), whereas a probe for the *Wolbachia* gene WD_0484 bound to a unique location (green, arrow) on chromosome 2L

Modularity

The Role of Translocation and Selection in the Emergence of Genetic Clusters and Modules

David Newth*
CSIRO Centre for Complex Systems Science
CSIRO Marine and Atmospheric Research
GPO Box 284
Canberra, ACT 2601
Australia
david.newth@csiro.au

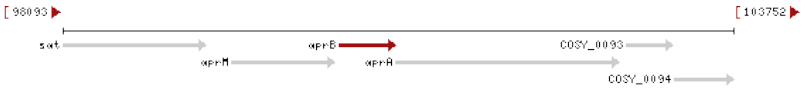
David G. Green
Faculty of Information Technology
Monash University
Clayton, VIC 3800
Australia
david.green@infotech.monash.edu.au

Abstract Biomolecular studies point increasingly to the importance of modularity in the organization of the genome. Processes such as the maintenance of metabolism are controlled by suites of genes that act as distinct, self-contained units, or *modules*. One effect is to promote stability of inherited characters. Despite the obvious importance of genetic modules, the mechanisms by which they form and persist are not understood. One clue is that functionally related genes tend to cluster together. Here we show that genetic translocation, recombination, and natural selection play a central role in this process. We distill the question of emerging genetic modularity into three simulation experiments that show: (1) a tendency, under natural selection, for essential genes to co-locate on the same chromosome and to settle in fixed loci; (2) that genes associated with a particular function tend to form functional clusters; and (3) that genes within a functional cluster tend to become arranged in transcription order. The results also imply that high proportions of junk DNA are essential to the process.

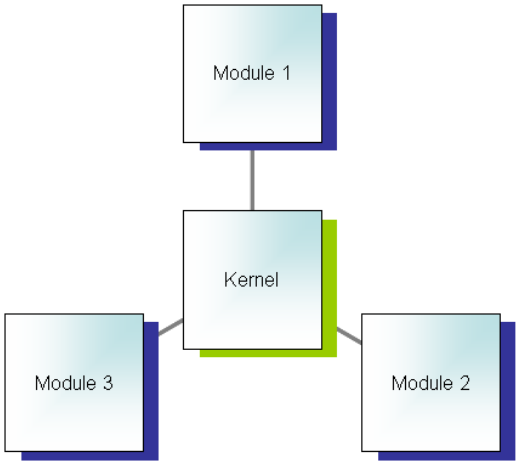
Keywords
Genetic modularity, self-organization, feedback, translocation, clustering

Genes tend to group in modules that facilitates the propagation of specific functions within a community of organisms

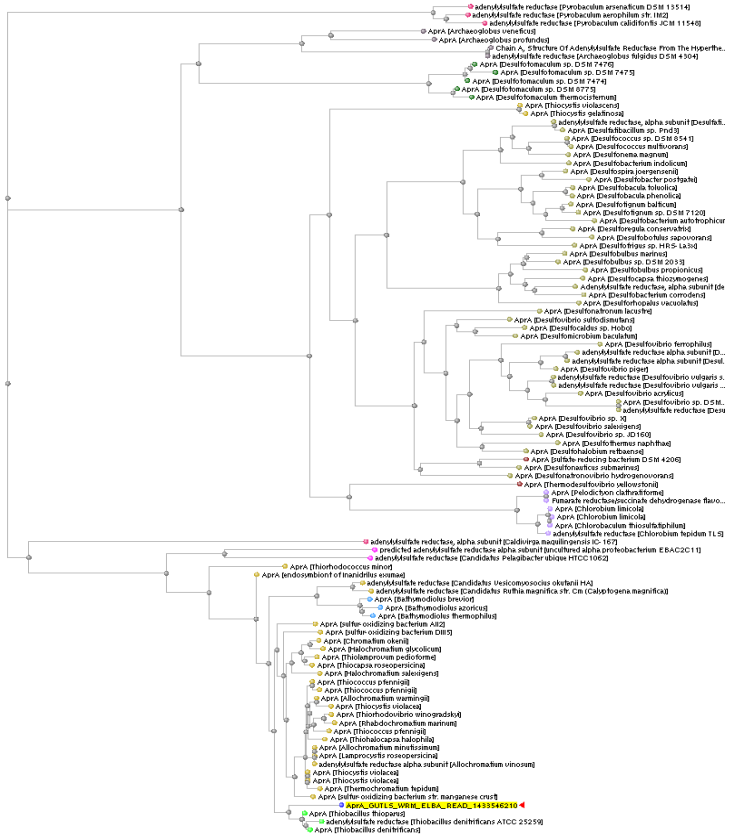
HGT and community dsr/sox metabolism in prokaryotes



sat-apr locus of *Vesicomyosocius okutanii*

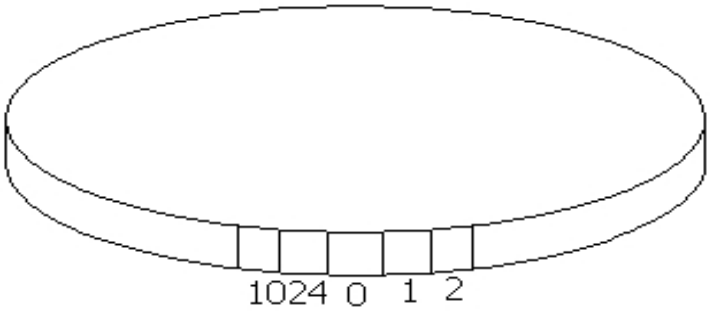


Artificial modular design. The kernel (housekeeping genes) and modules involved into process of adaptation towards a particular environment [Kuznetsov, 2008]



The AprA tree was produced using *Alvinella pompejana* symbiont reference and NCBI BLAST pairwise alignments

A recursive function in agents



Each creature has a circular genome consisting of 1024 ‘genes’, only one of them is active and coded by color with mod(1024)

$$\mathbf{T(i+1) = |[T(i) + P(i)] / 2 * R|_{mod(1024)}}$$

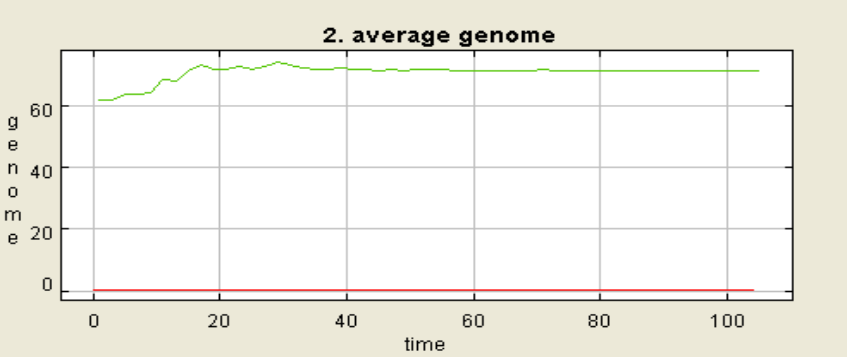
$$\mathbf{P(i+1) = T(i+1),}$$

where **T(i)** is the color code of the individual Spermatozoon and **P(i)** is the color code of the individual Ovum at the time **i** of breeding. **R** is the mutation parameter on the interval]0, 4]

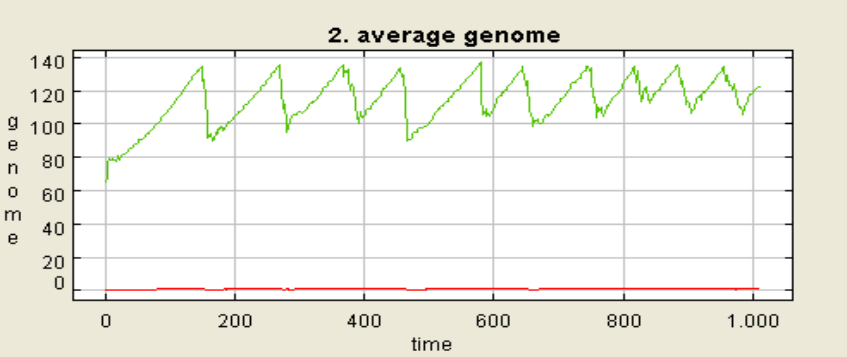
Viruses, primordial sperm cells, parasites and symbionts, or whatever [Kuznetsov, 2004]

Complexity can be taken easy

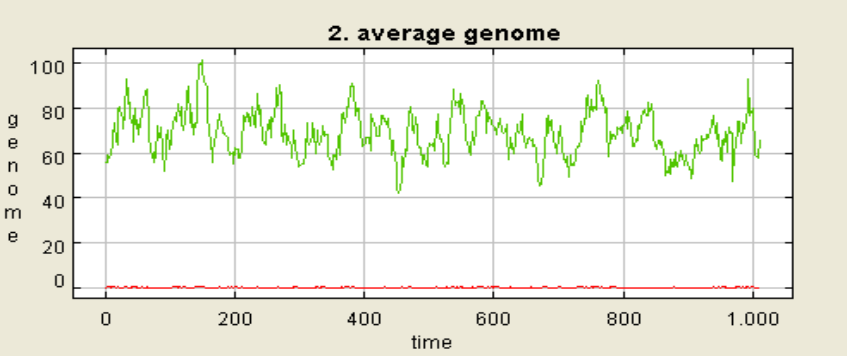
The system demonstrated ordered (R<=1) and complex (R>1) regimes



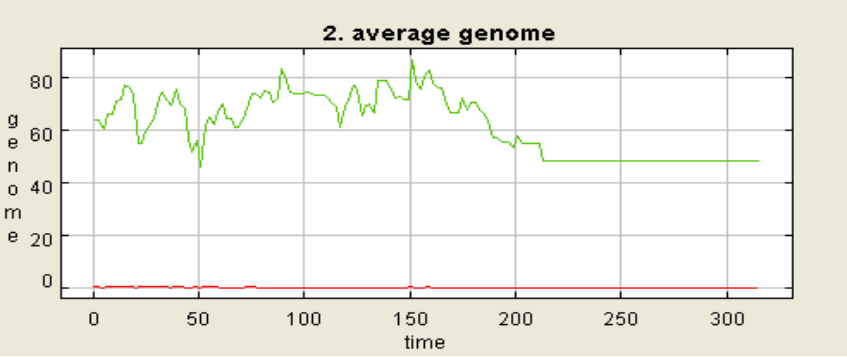
stable focus, R=1



periodic, R=1.01

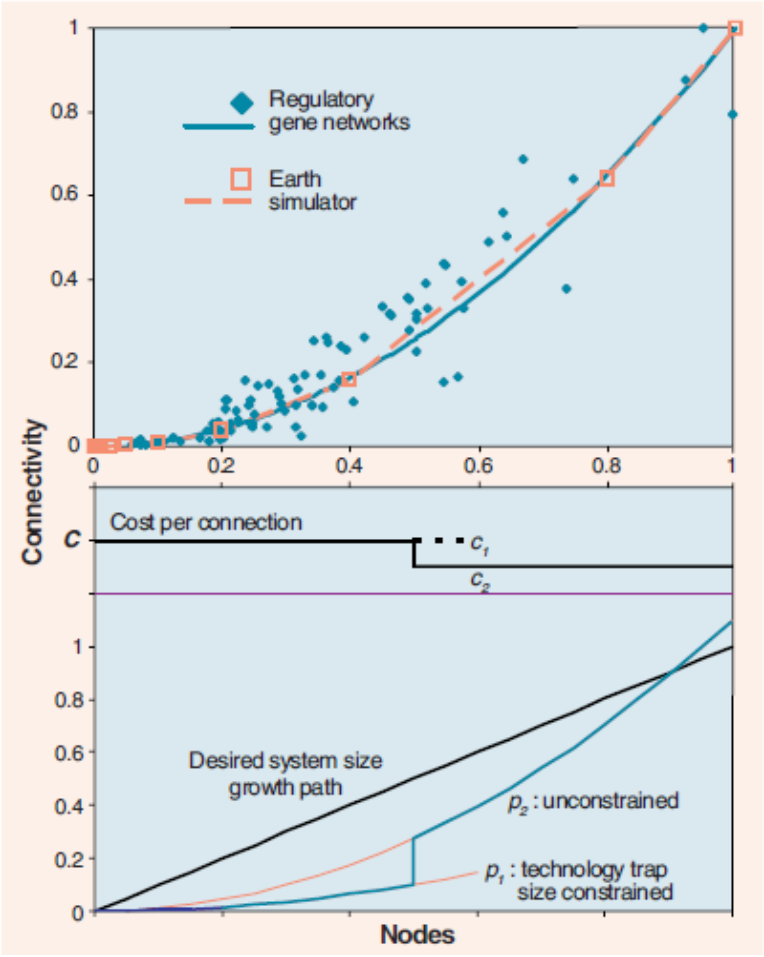


chaotic, R=3



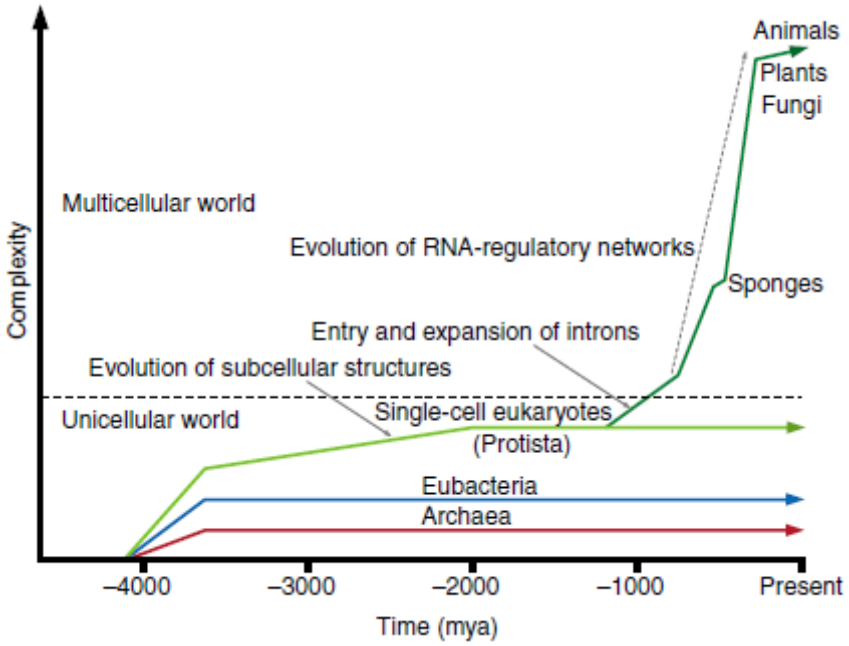
strange attractor, R=4

Graph connectivity and complexity



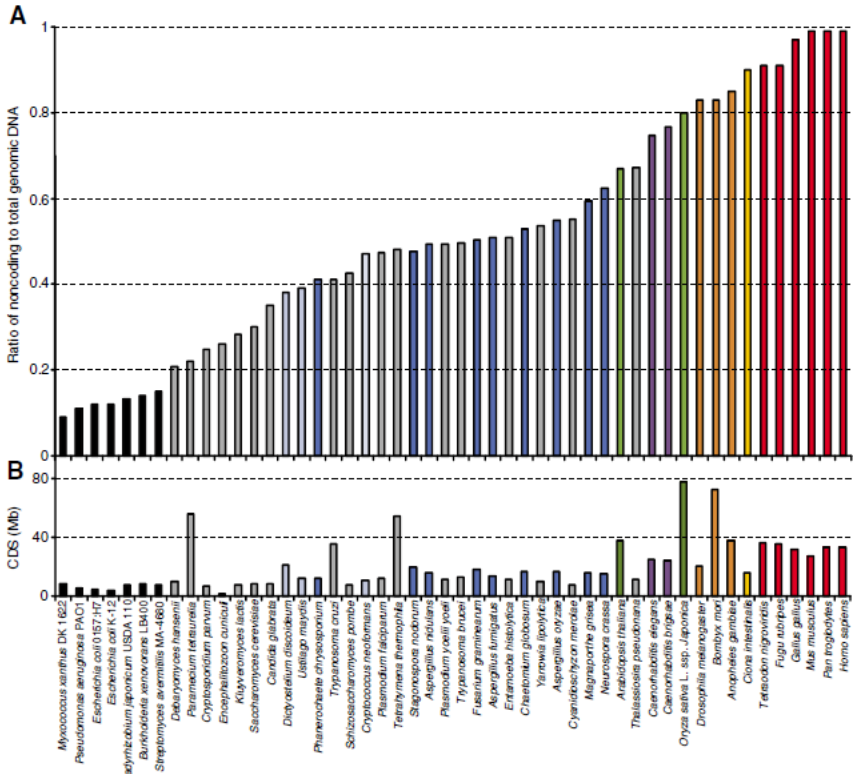
[Mattick, Gagen, 2005]

A number of connections in a net goes up as a square of the number of nodes. If the informational resources were restricted, the growth is slow down. The system needs to invent a new kind of regulation



A simplified view of the biological history of the Earth [Mattick, 2007]

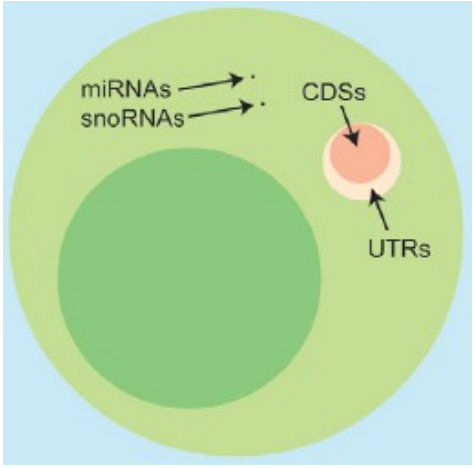
The G-value paradox



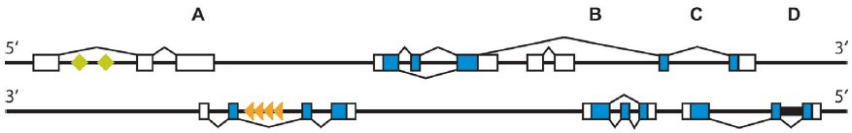
[Taft et al, 2007]

- Human (and other vertebrates) have approximately the same number of protein-coding genes as *C.elegans* (30,000-40,000 vs. 20,000), and less than those of some plants (rice ~40,000). Most of the proteins have similar functions in human and nematode
- **Where is the information that programs human complexity?**
- Additional questions
  - Bottle neck effect
  - Red Queen effect
  - Cambrian explosion, role of transpositions

Modern RNA world

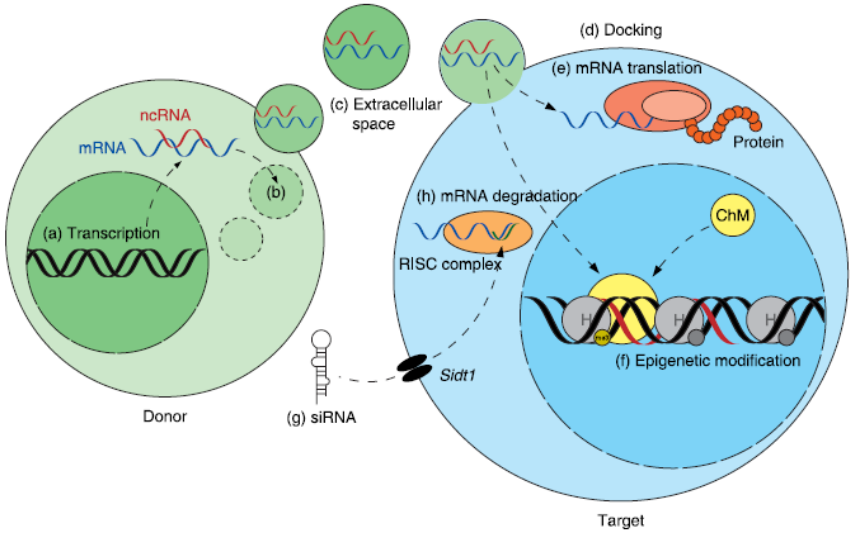


Transcription in mammals. The area of the box represents the genome. The large green circle is equivalent to the transcriptom, with the darker green area corresponding to transcripts from both strands. CDSs are protein-coding sequences, and UTRs are untranslated regions in mRNAs. The dots indicate (and in fact overstate) the proportion of the genome occupied by known miRNAs and snoRNAs [Mattick, Makunin, 2006]



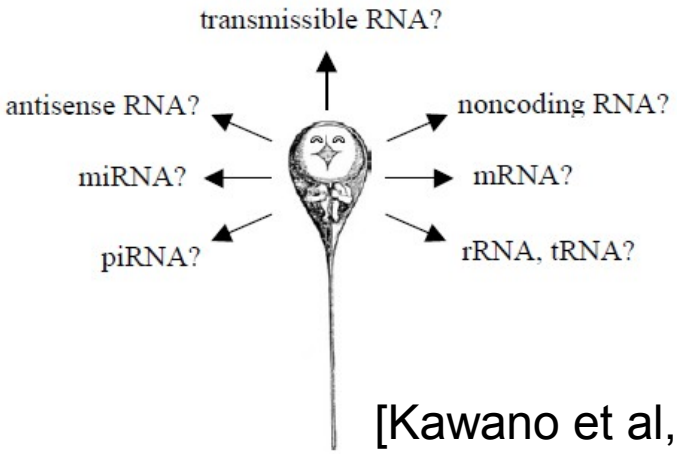
Complexity of the transcriptional landscape. White boxes represent non-coding exonic sequences and blue boxes protein-coding exonic sequences. Green diamonds represent snoRNAs and orange triangles represent miRNAs. Indicated are (**A**) antisense transcripts with overlapping exons, (**B**) nested transcripts on both strands, (**C**) antisense transcripts with interlacing exons and (**D**) retained introns [Mattick, Makunin, 2006]

RNAs as extracellular signalling molecules

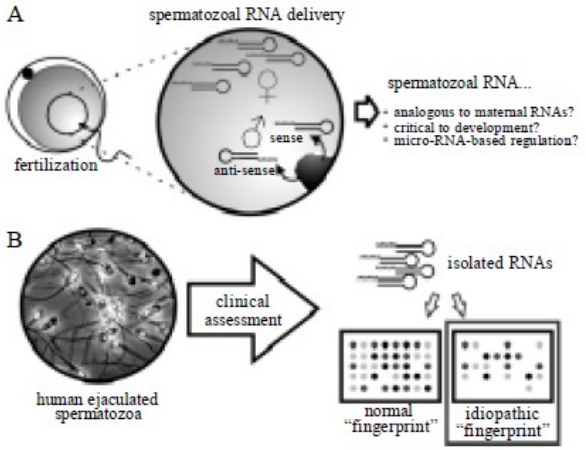


- Known and putative signalling mRNA (blue) and ncRNA (red) are (**a**) transcribed in the donor cell. These RNAs are then (**b**) trafficked and packaged into vesicles, which are emitted into (**c**) the extracellular environment. (**d**) The vesicles then dock and fuse with the target cell, releasing their RNA content. The mRNA may then be (**e**) translated in the target cell and the ncRNA may guide the chromatin modifying proteins (ChM) to establish (**f**) the new epigenetic state
- In addition (**g**), extracellular RNA molecules, such as siRNAs or miRNAs, may be transferred across the plasma membrane by specific receptors and channels, such as Sid-1. (**h**) These RNAs may regulate the inhibition of translation or mRNA degradation in the cell [Dinger et al, 2008]

‘Unknown’ sperm cells, current research

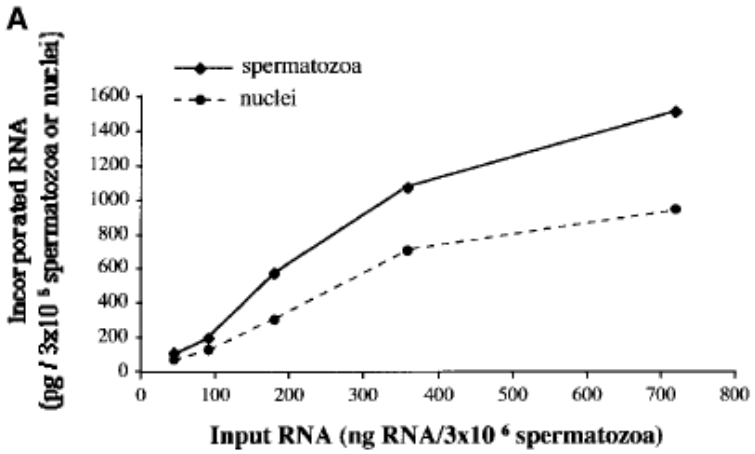


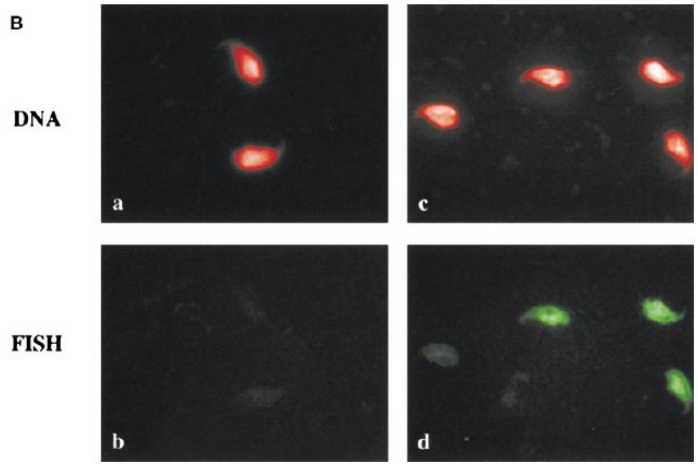
[Kawano et al, 2007]



- **Do spermatozoa much more than a simply delivery of male genome at the fertilization?**
- (**A**) Sperm cells transfer not only their haploid genome into the embryo, but also many RNA transcripts. Among these are transcripts involved in early development, as well as miRNAs that may play a role in the embryo regulation. (**B**) The majority of male factor infertility is unknown, as sperm appear normal. RNA isolated from spermatozoa obtained from infertile men can be used to clinically assess putative genetic or environmental disturbances. This will not only explain genetic processes of sperm differentiation and early embryo development, but also provide a therapy [Martins, Krawetz, 2005]

Evidence of RNA-uptake by mouse spermatozoa

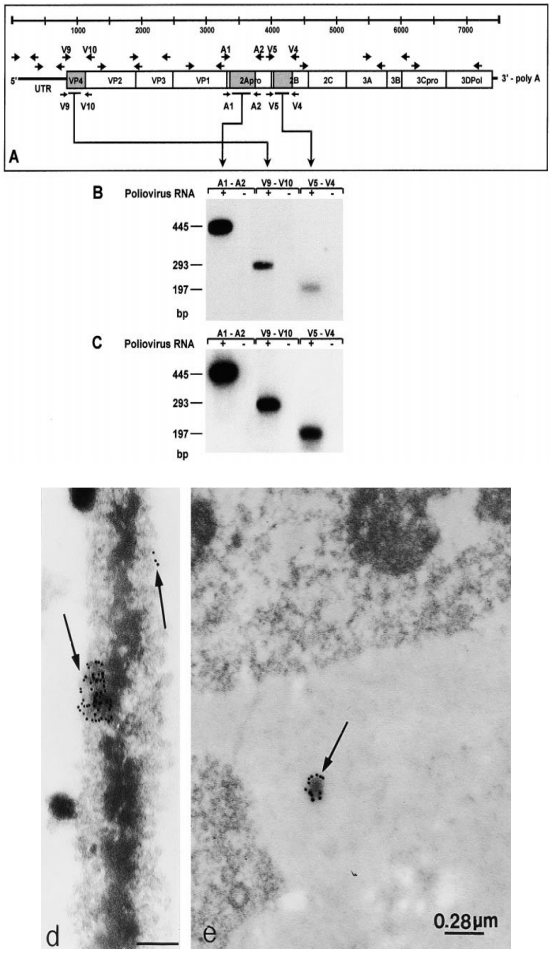




- **Association of the radioactive end-labelled poliovirus RNA with mouse epididymal sperm cells**. Spermatozoa were incubated with poliovirus RNA. After 30 min, spermatozoa from each mixture were washed and divided in 2 aliquots. The first one was dissolved in scintillation cocktail and counted to measure the RNA uptake; the second aliquot was used for nuclei purification, then treated and counted as for whole cells to measure nuclear internalization

- **FISH of poliovirus RNA in isolated nuclei from RNA-treated spermatozoa**. (**a** and **c**) Sperm nuclei stained with DAPI and pseudocolored in red; (**b**) FISH with nuclei from spermatozoa incubated with buffer only; (**d**) FISH with nuclei from spermatozoa incubated with poliovirus RNA. The signal associated with the biotinylated probe is pseudocolored in green [Giordano et al, 2000]

Reverse transcriptase activity was found in mouse sperm cells



- Human poliovirus RNA replicates through a RNA (–) strand with no DNA intermediate. The poliovirus RNA was taken up by sperm cells, then reversetranscribed to cDNA copies, which were transferred to oocytes during IVF and further transmitted to 2-cell embryos

- **PCR of cDNA copies in sperm cells and 2-cell embryos after incubation with poliovirus RNA**. (**A**) Map of poliovirus RNA chromosome, (**B**) spermatozoa, and (**C**) 2-cell embryos. Amplified cDNAs were visualized by hybridization with internal oligonucleotide probes increasing specificity
- **Immunoelectronmicroscopy with anti-RT antibody showed that RT molecules were associated with the sperm nuclear scaffold**. (**d**) Sperm nuclear scaffolds, (**e**) HIV-infected T-lymphocyte [Giordano et al, 2000]

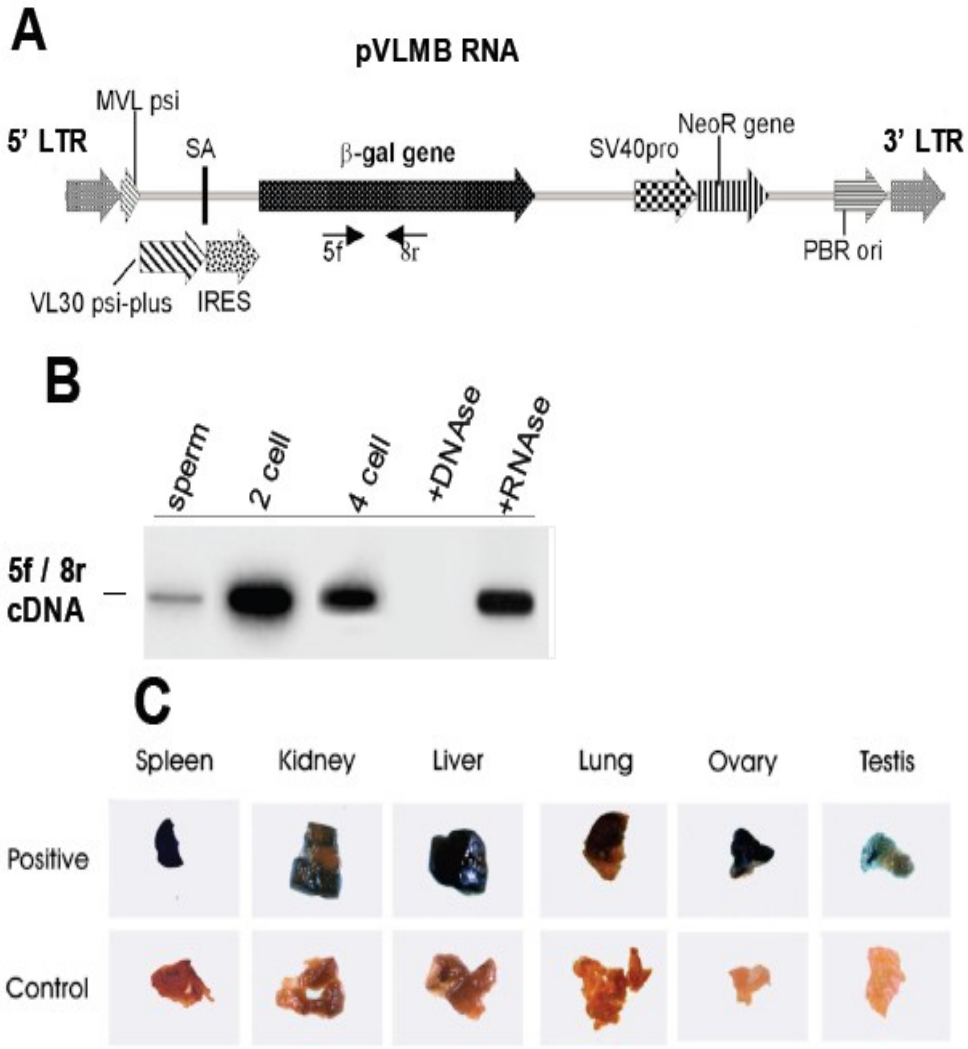
Attention

- The endogenous RT of mammalian cells is encoded by 2 large families of retrotransposons: **1)** LINE-1 (Long Interspersed Nuclear Elements), accounting for about 17% of the human genome, and **2)** endogenous retroviruses (ERVs), about 8% of the genome. Retrotransposable elements together comprise much as 45% of the human genome, while the protein-encoding genes represent a mere 1.2% [Deininger et al, 2003]

- RT-encoding genes are constitutively expressed at detectable, albeit variable, levels in several normal tissues [Seifarth et al, 2004]. Preferential locations of expression of retroviral/retrotransposon genes are the mammalian genital tract [Kiessling, 1984], male germ cell precursors [Branciforte and Martin, 1994] and gametes [Kattstrom et al, 1989]

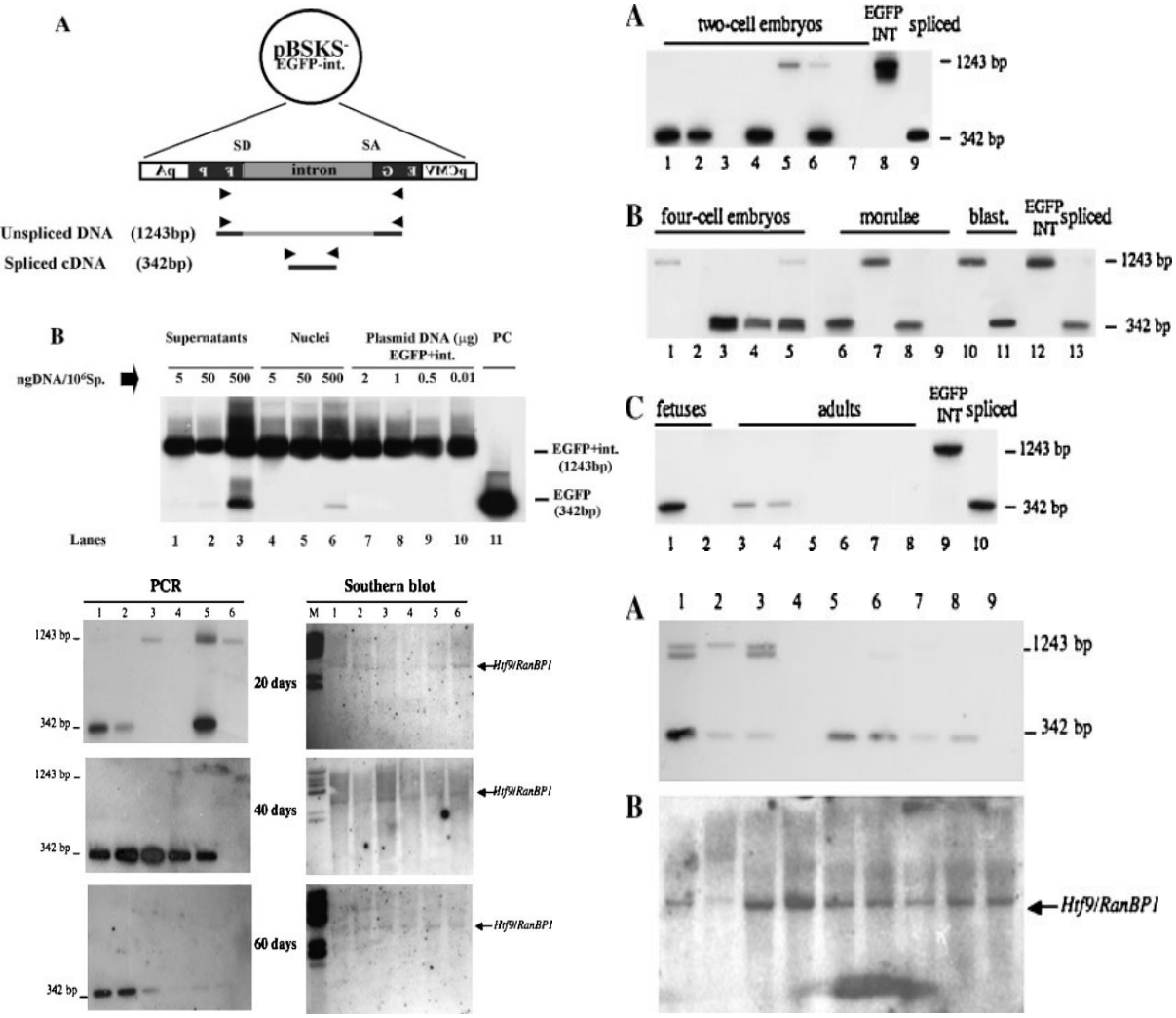
- Nevirapine, an inhibitor of RT, when present soon after insemination, caused an irreversible arrest of development at the 2- to 4-cell stages. RT inhibition caused a substantial reprogramming of the gene expression profile in arrested embryos, involving both developmentally regulated and constitutively expressed genes [Pittoggi et al, 2003]

Sperm RT as a mediator of new genetic information



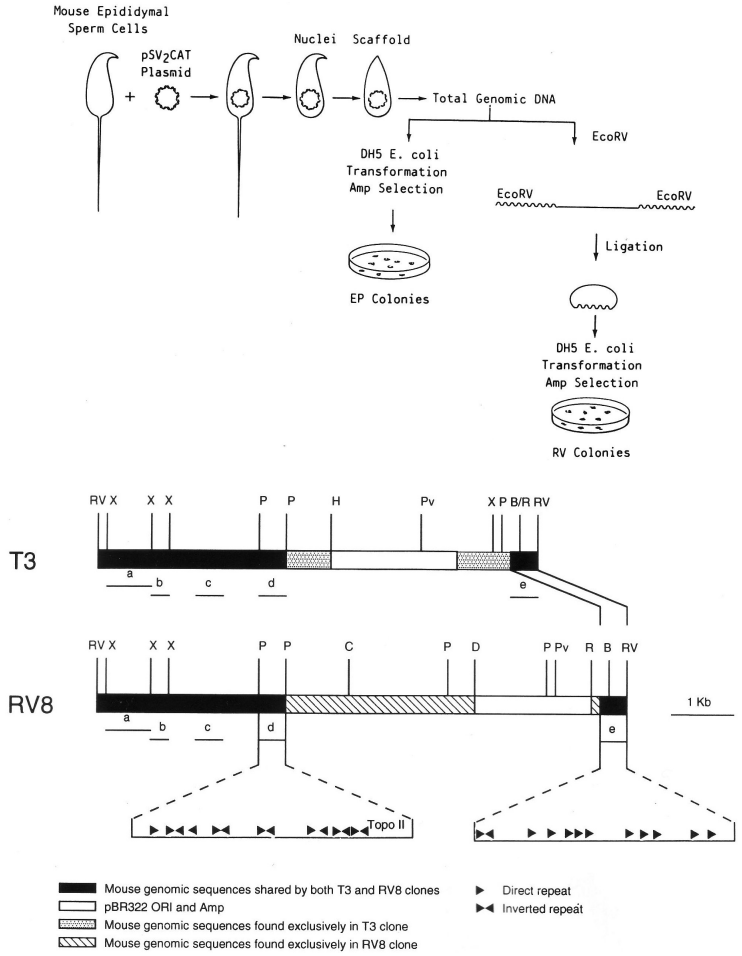
Sperm cells were pre-incubated with RNA transcribed from a β-gal reporter gene-containing construct (**A**) and then used in IVF assays to generate 2- and 4-cell embryos. The β-gal cDNA molecules were amplified by direct PCR in spermatozoa and in embryos at both cell stages (**B**). 2-cell embryos were then implanted into foster mothers to obtain F0 individuals and F1 mouse progeny. Animals were randomly selected from both F0 and F1 populations and the DNA extracted from various organs was screened for the presence of β-gal cDNA copies by direct PCR. The results indicated that β-gal cDNAs were: 1) present in low copy number (<1/cell); 2) mosaic distributed in tissues of adult animals; 3) sexually transmitted from founders to F1 progeny in a nonmendelian way, and 4) again propagated as mosaic, low-copy number sequences in F1 tissues. (**C**) X-gal staining [Sciamanna et al, 2003]

Generation of retro-genes upon interaction of mouse spermatozoa with exogenous DNA



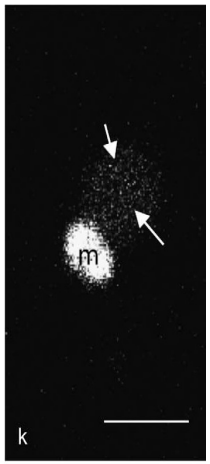
The EGFP gene under CMVe promoter interrupted by *γ*-globin intron insertion was used. Mouse sperm cells were incubated with the pBSKS-EGFP-INT. Total DNA was extracted and amplified from supernatant and nuclei of sperm cells. The spliced EGFP was found in supernatant and nuclei. This experiment demonstrated that the original DNA was transcribed in RNA, processed and then was reversetranscribed into cDNA. After fertilization in vitro, the cDNA was revealed by PCR in F0 pre-implantation embryos, fetuses and adult mice (unspliced copies were observed rarely). A decreasing amount of EGFP-containing cDNAs in founder animals during development was observed. The authors were able to demonstrate the cDNA propagation in various tissues of founder animals by PCR (the unspliced fragments were revealed with rearrangements). Transmission of EGFP cDNA to F1 offsprings was showed by PCR (20-88%). An expression of EGFP protein in the vascular epithelium was reported that correlated with the tropism of CMVe promoter [Pittoggi et al, 2006]

Integration of the pSV2cat into the unique site of mouse sperm genome

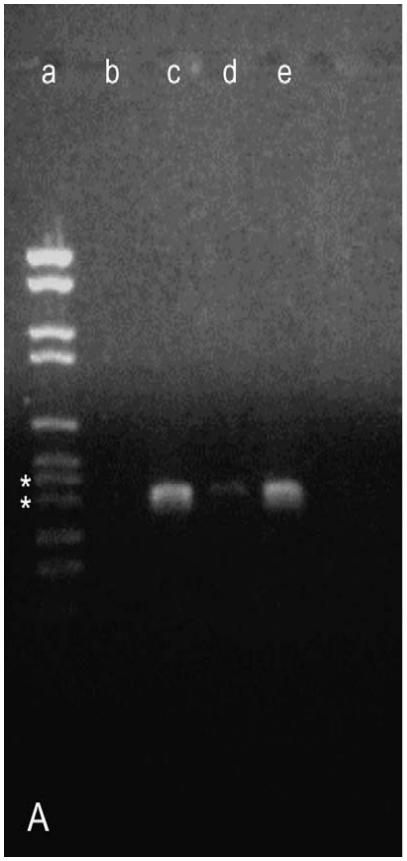


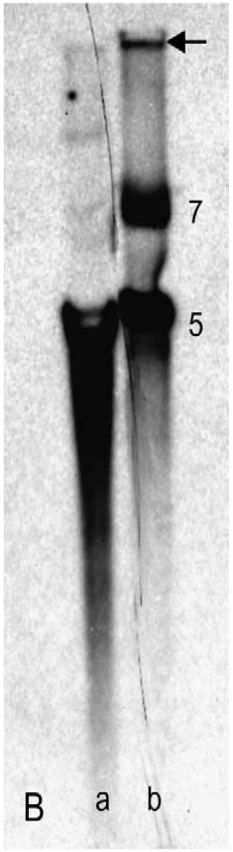
- pSV2cat is rescued from mouse sperm genome after sperm/DNA incubation (EP – episome, RV – EcoRV digestion and ligation)
- 2 mouse DNA sequences, identical in T3 and RV8 clones (solid), flank the unidentified DNA fragements (dotted, hatched) within which the plasmid (open) has been integrated (Amp^R and *ori* of pSV2*cat*: 707-2542 in T3, 751-2620 in RV8). The TopoII consensus sequence adjacent to the one terminus of the site of integration; **b**,**c**,**e** – Alu-like repeats
- <u>The same (solid) sequences were found in 14 randomly selected clones by DNA hybridization</u> [Zoraqi, Spadafora, 1977]

Transfection of spermatozoa in bivalve molluscs using naked DNA



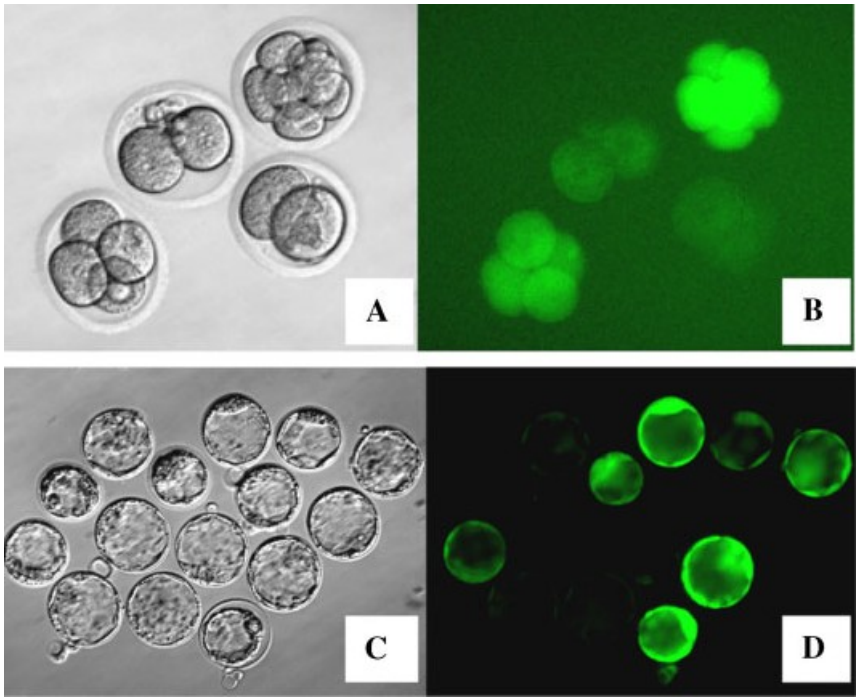
Mytillus galloprovincialis spermatozoon transfected with the rhodamine-labeled pGeneGrip-GFP construct. Spermatozoon observed under fluorescence. The nuclei (arrows) contain numerous fluorescent dots; m, mitochondrial groups, which are deeply fluorescent. Bar 2.5 µm [Guerra et al, 2005]





- (**A**) PCR of the 411-bp GFP-segment of the pGeneGrip-GFP in DNA purified from *M.galloprovinciallis* sperm cells. Lane **a**, molecular weight marker. The two asterisks show the bands of 453 and 394 bp. Lane **b**, DNA from control spermatozoa. Lanes **c** and **e**, the amplified 411-bp bands in DNA from treated spermatozoa. Lane **d**, positive control (10 ng plasmid)
- (**B**) Southern blot of the 5757 bp pGeneGrip-GFP (lane **a**) and DNA from sperm cells incubated with plasmid pGeneGrip-GFP (lane **b**). DNA was digested with EcoRI. In lane **b** two bands, one of 7 kb and the other of more than 10 kb (arrow) are observed

DMSO-SMGT in mouse



- **The mouse embryos expressed EGFP on the different stages of development after DMSO-SMGT and fertilization *in vitro***
- **A** and **C** show the developed embryos at different stages, **B** and **D** represent the green embryos under fluorescence microscope. In picture **D**, some blastocysts are not ‘transgenic’ ones, at meanwhile the green fluorescence cannot be detected [Shen at al, 2006]

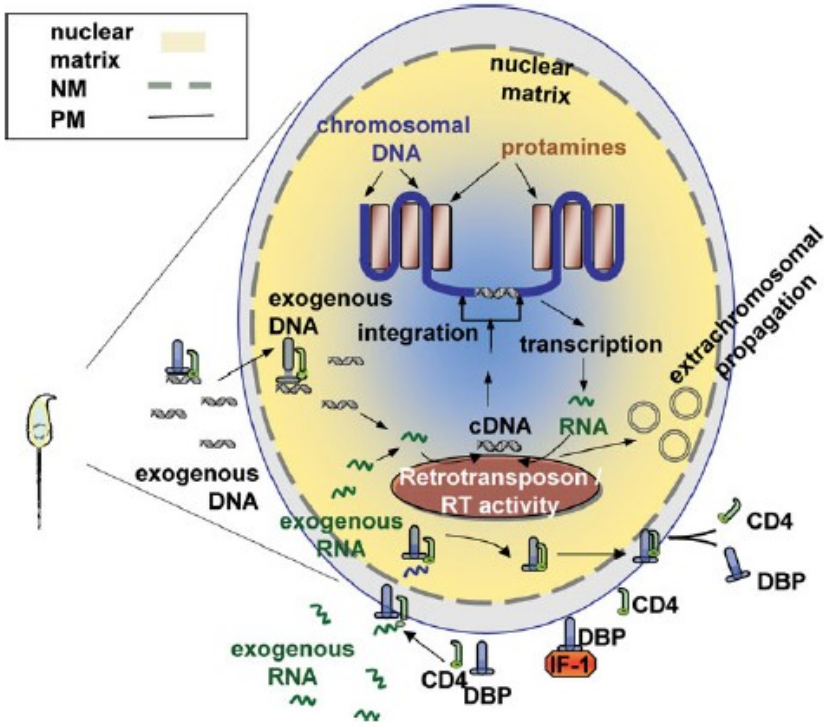
Expression of CMV-*lacZ* in *Misgurnus fossilis* fry after electro-SMGT

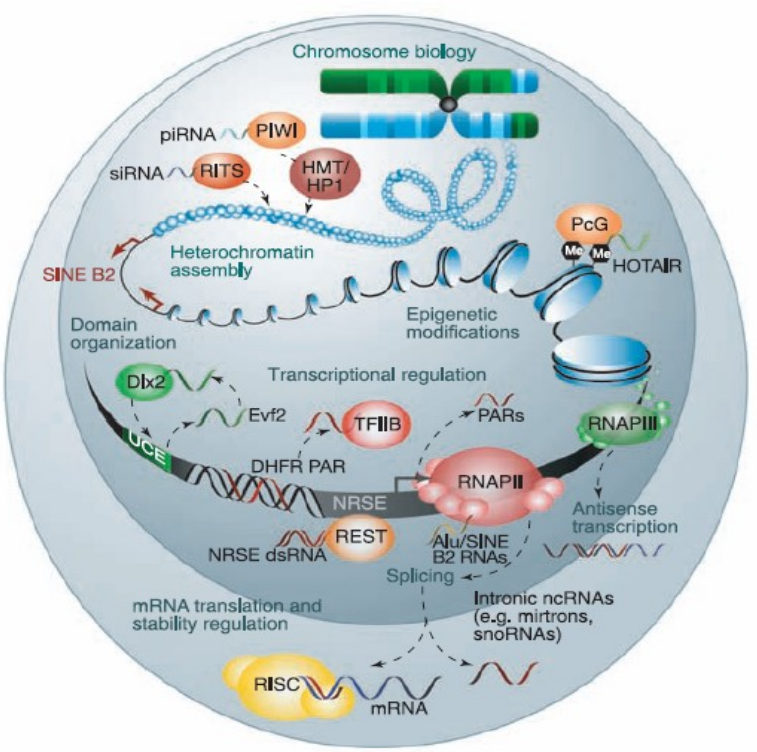
Control loach fry – mock analysis
Experimental β-gal-positive fry 72 h after the eggs fertilization by sperm cells transfected with pcDNA3-*lacZ*

Only blue spots, dots and dashes are probably a result of ‘transgene’ elimination during development

Sperm cells were squeezed out of the loach testis and intensively washed. Electric discharge (V=150 V, R=150 Ω, C=20 μF) was passed through the cell’s suspension containing of 0.5 μg/ml DNA. Transfected sperm was added to eggs for fertilization. After development the embryos were fixed by 2.5% glutaraldehyde and stained by X-gal [Andreeva et al, 2003]

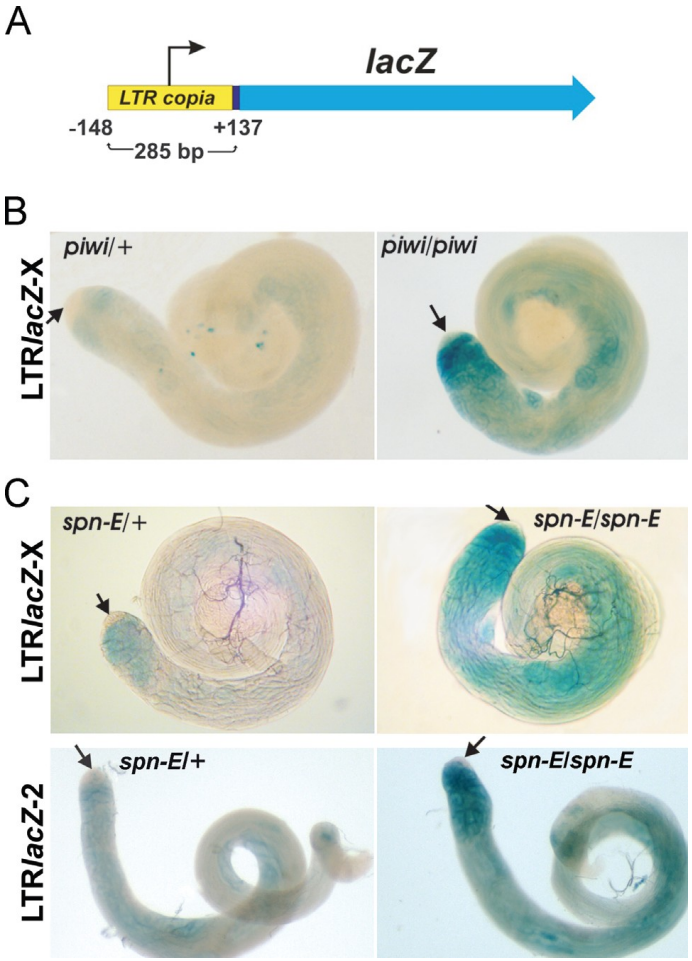
Summary: SMGT and RNAi





- *Spadafora C.* Sperm-mediated 'reverse' gene transfer: a role of reverse transcriptase in the generation of new genetic information. Hum Reprod. 2008 Apr;23(4):735-40
- *Amaral PP, Dinger ME, Mercer TR, Mattick JS.* The eukaryotic genome as an RNA machine. Science. 2008 Mar 28;319(5871):1787-9

PIWI-control of *copia* LTR in *Drosophila* male germline

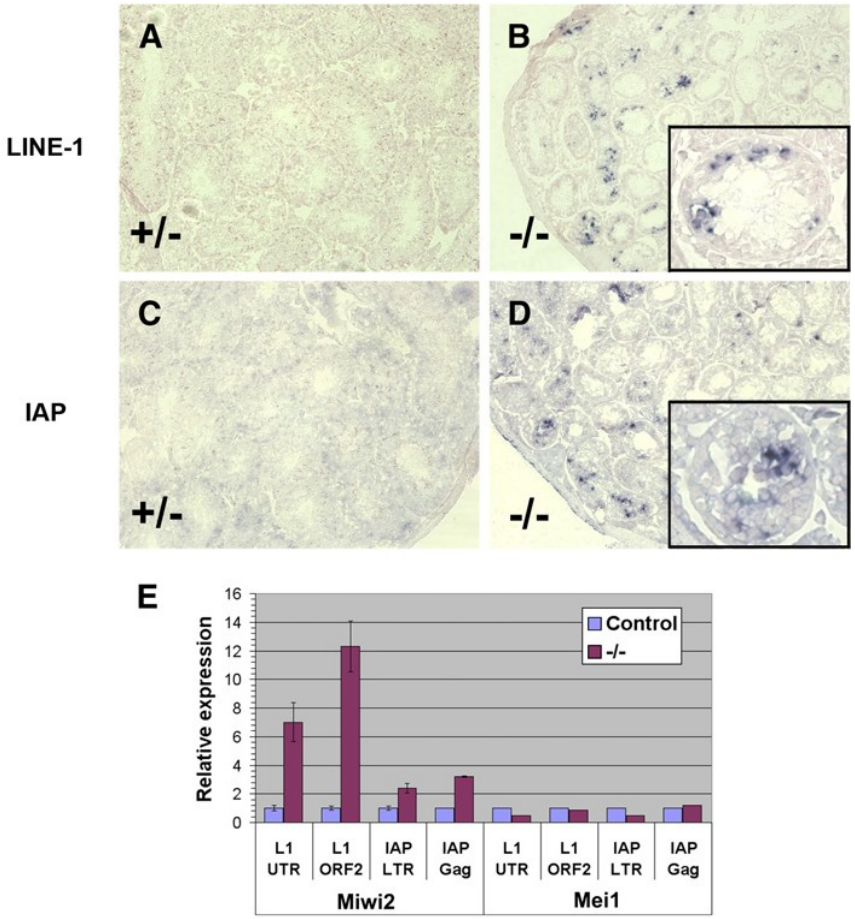


- RNAi plays a role in defence against transposable elements and viruses. Protein of the Argonaute family prevents retrotranspositions. PIWI is involved in silencing of LTR retrotransposons in testes. The disturbance of RNA silencing system in germinal cells might cause a transposition burst

- **Overexpression of the copiaLTR-lacZ in piwi2 and spn-E1 testes.** (**A**) The copiaLTR-lacZ construct comprises full-size copia LTR fused to the lacZ reporter. X-gal staining of testes (**B**) from piwi2 males carrying the copiaLTR-lacZ on the X-chromosome, (**C**) from spn-E1 males carrying the copiaLTR-lacZ on the X (upper panels) and second chromosomes (lower panels)

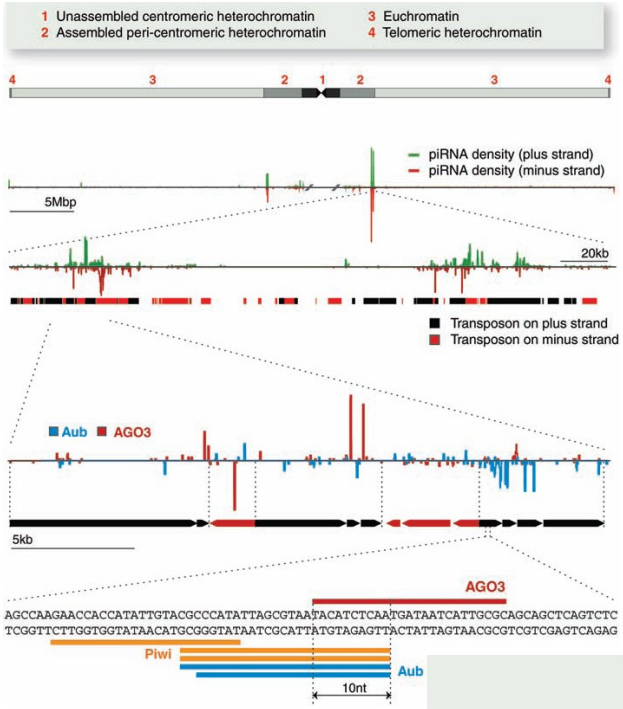
- A level of lacZ expression is greatly increased in the homozygous piwi/piwi and spn-E/spn-E testes as compared with heterozygous piwi/+ and spn-E/+. Arrows indicate the apical tips of testes, where β-galactosidase activity is not detected [Kalmykova et al, 2005]

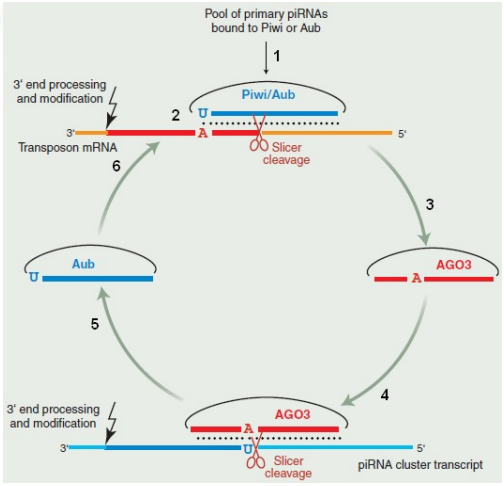
MIWI repression of LINE-1 and IAP in the mouse male germline



- MIWI2 is a mouse Piwi-family member. piRNAs become abundant in germ cells around the pachytene stage of prophase of meiosis I. Disruption of MIWI (-/-) interfered with transposon silencing in the male germline
- (**A–D**) **In situ hybridization of testes** of the indicated genotypes with probes recognizing the sense strands of LINE-1 and IAP (Intracisternal A particles) elements
- (**E**) **Quantitative RT-PCR analysis** of transposable elements in 14-day-old animals [Carmell et al, 2007]

Adaptive defence in the transposon arms race



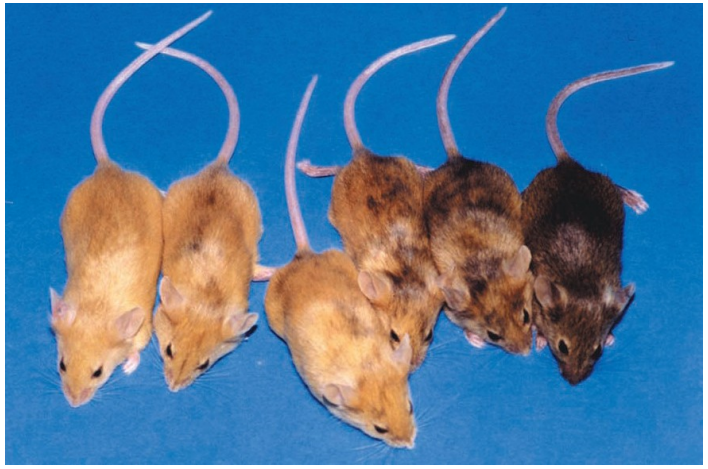


The genomic origin of piRNAs. Most piRNA clusters in *Drosophila* are located in pericentromeric or telomeric heterochromatin. The piRNA clusters range from several to hundreds of Kb. They are devoid of protein coding genes and instead are highly enriched in transposons and other repeats, which occur in the form of nested, truncated, or damaged copies that are not capable of autonomous expression or mobilization

Ping-Pong model of piRNA biogenesis in *Drosophila*. The Ping-Pong amplification loop begins with a transposon-rich piRNA cluster giving rise to a variety of piRNAs (**1**). When piRNA binds a complementary target (a transposon mRNA), Piwi/Aub complexes cleave the duplex 10 nt from the 5′ end of piRNA. This not only inactivates the target but also creates the 5′ end of new AGO3-associated piRNA (**2**). Loaded AGO3 complexes are also capable of cleaving complementary piRNA cluster transcript (**3**). The cleavage of cluster transcripts by AGO3 (**4**) then generates additional copies of the original antisense piRNA, which would enter Aub (**5**) and become available to silence active transposons (**6**). The combination of these steps can form a self-amplifying loop. Signatures of this amplification loop are apparent in mammalian prepachytene piRNAs [Aravin et al, 2007]

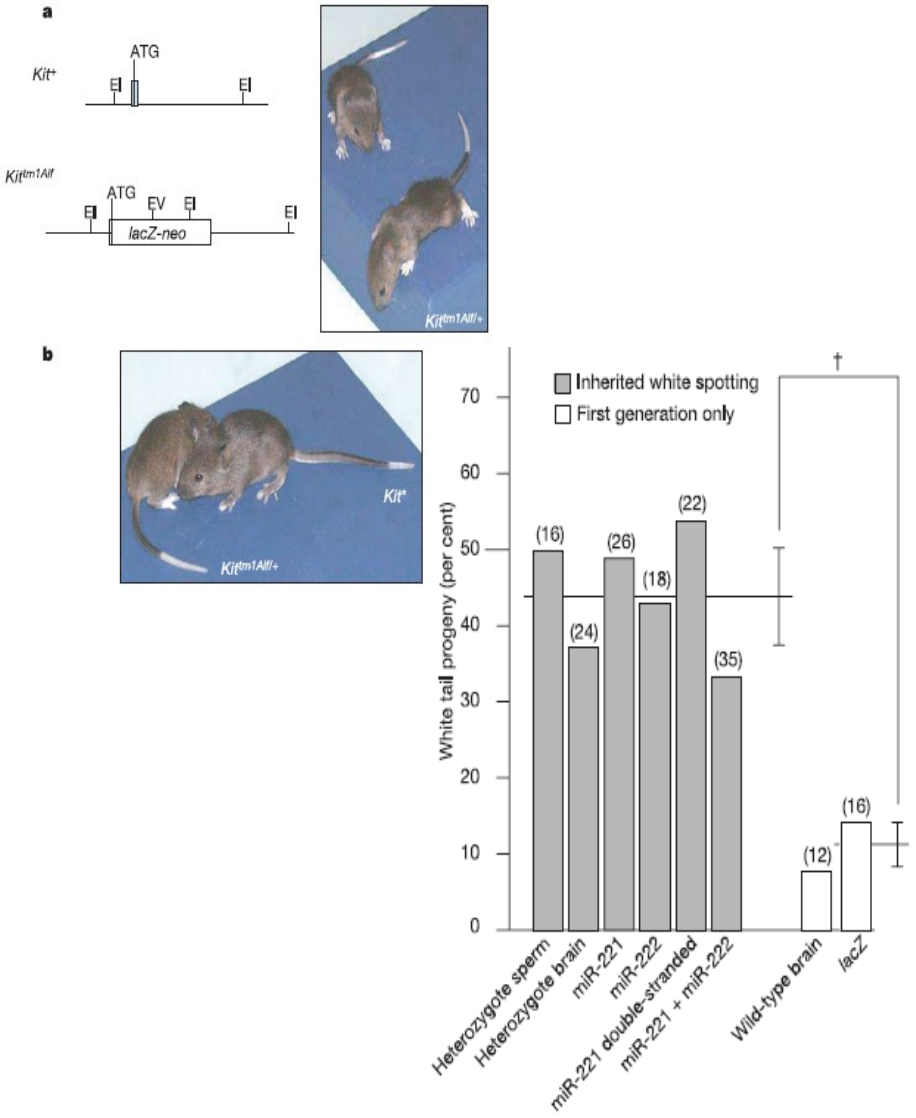
The stochastic nature of retrotransposon activity

- **Retrotransposons are silenced by RNA-mediated mechanism**
- **Stochastic activation and silencing of retrotransposons during early embryogenesis leads to individuals with different phenotypes**



- Isogenic yellow agouti (A^{vy}) mice display a range of phenotypes, because an IAP retrotransposon has inserted 100 kb upstream from the agouti hair cycle specific promoter
  - When the IAP is active, the agouti protein (*A*) is transcribed from a cryptic promoter in the LTR of the IAP, the normal program of *A* expression is distorted. Ectopic expression of *A* results in yellow fur, as well as diabetes and tumours (**left**)
  - When the IAP is silent, the agouti protein is expressed specifically and the fur is agouti (**right**)
- Mice in which the IAP is completely silent are phenotypically indistinguishable from agouti mice [Whitelaw, Martin, 2001]

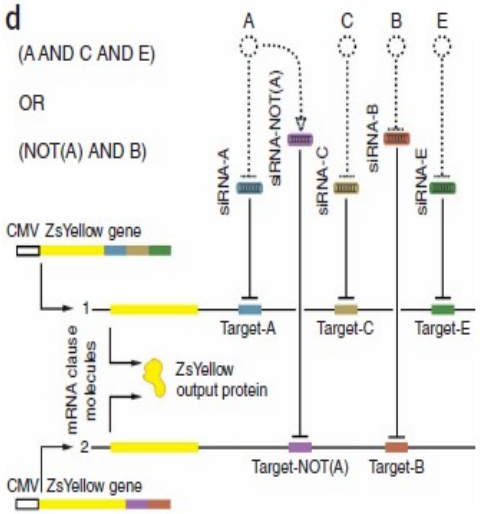
RNA-mediated inheritance



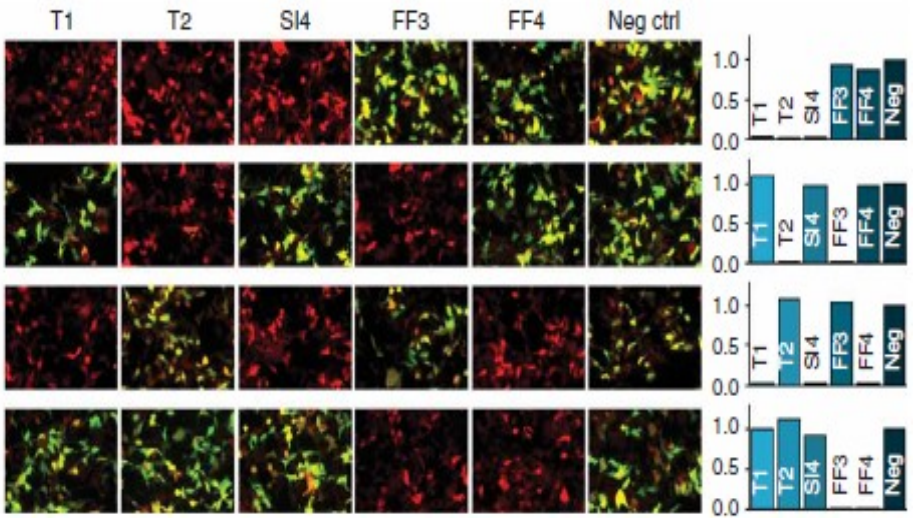
- **Mutation, caused by *LacZ* insertion in the *Kit* gene**, generates white patches on the tail of heterozygotes due to the reduced expression of Kit tyrosine kinase receptor (**a**). Some progenies of those mice, who inherited 2 wild-type *Kit* alleles, still exhibit the white patches (**b**)

- **The mutant phenotype is associated with accumulation of abnormal RNA molecules in spermatozoa**. Microinjection of this RNA in zygotes induced a heritable white tail phenotype (**c**) [Rassoulzadegan et al, 2006]

RNAi evaluator of DNF and CNF logic in human embryonic kidney cells

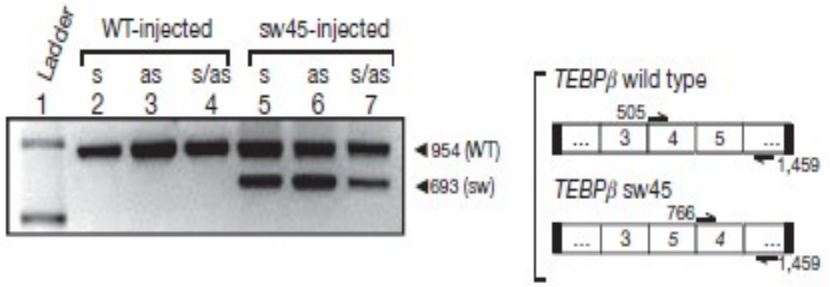


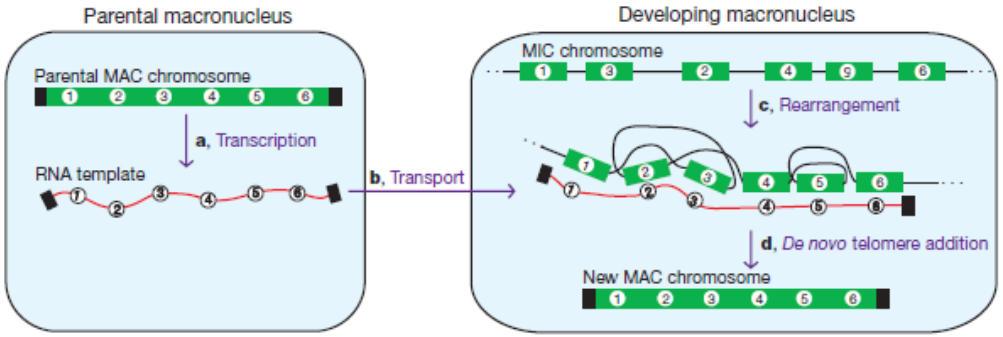
- **Design of a disjunctive normal form (DNF)** where dotted lines are input linked to mediator siRNAs that in turn target the 3'-UTRs of the ZsYellow RNAs (**d**)
- **The expression after 48 h** demonstrates a 16-fold difference between the False and True outputs. Low levels of ZsYellow protein expression (red), coexpression of both proteins (green and yellow) [Rinaudo et al, 2007]



RNA-guided recombination may be widespread

Oxytricha trifallax is a unicellular eukaryote with 2 nuclei: germline micronucleus and somatic macronucleus. Micronucleus is transcriptionally inert, but transmits the germline genome to next generation. Macronucleus provides gene expression, but degrades after fertilization, when new micro- and macronucleus develop. Differentiation in ciliates involves deletion of transposons and chromosome fragmentation into 2 kb 'nanochromosomes' with just 1 gene, accomplishing 95% genome reduction and compressing 1 Gb germline (30,000 genes) into 50 Mb. It is an accurate **mechanism of DNA rearrangements by maternal RNA cache**

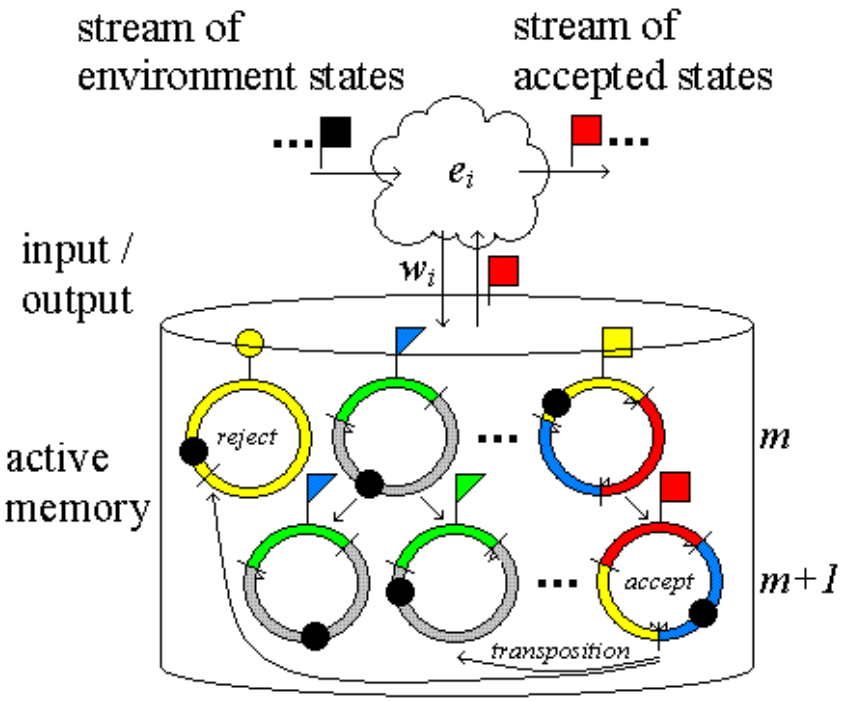




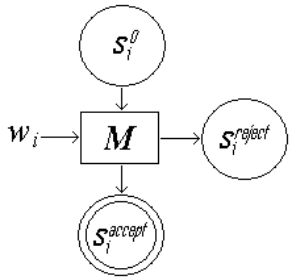
- **RNA template microinjection of Telomer-End-Binding Protein subunit β**. Sense (s), antisense (as), and combined (s/as) RNA were microinjected in both wild-type and switched orientations. Lanes 5–7 display the segments 4 and 5 have been switched (lower band)

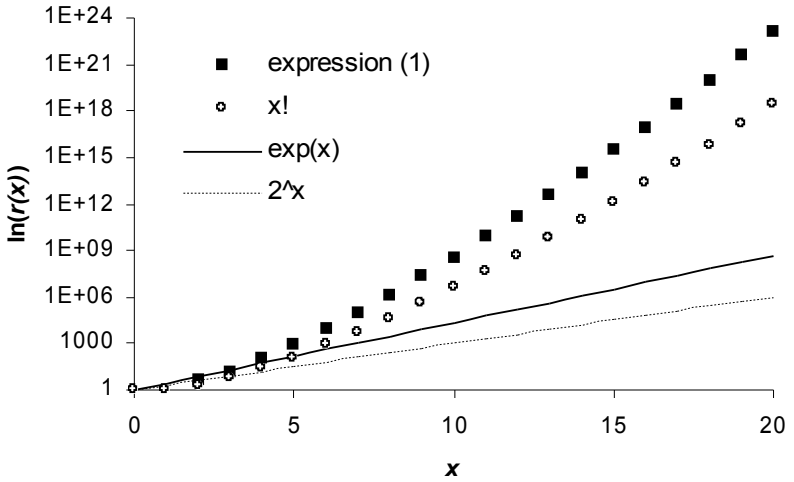
- **RNA-guided DNA rearrangements during macronuclear development**. **a**, RNA transcription of all DNA in the old macronucleus before its degradation. **b**, Transport of these RNA to the newly developing macronucleus, where they may act as scaffolds to guide DNA rearrangements (**c**). **d**, Telomere addition (black rectangles) and amplification of new macronuclear nanochromosomes [Nowacki et al, 2007]

Argo-machine



$$\begin{cases} r_0 = r_1 = 0, \\ r_x = 2^x * (x-1)!, \; x \geq 2 \end{cases} \quad (1)$$





The system operates on sequence inputs and memory, uploads the memory and yields outputs [Kuznetsov et al, 2006]

Conclusion

- Spermatozoon is not only a vector for the male haploid genome, but also the carrier of an additional information in the form of RNA/DNA molecules
- Occasional templating leads to different phenotypes. “There is no reason to suppose that each case of inheritance follows Mendelian rules”
- Information transfer from RNA to DNA involves (1) reverse transcription and (2) RNA-guided DNA rearrangements
- The consequent rounds of direct and reverse transcriptions are an important source of biological complexity
- It is a very big RNA world, the combinatorial power of the RNA-machine is incredible

Not the end of the story: RNA-mediated subtle genetic flow

Thanks to

- **Corrado Spadafora**, Istituto Superiore di Sanita, Rome, Italy
- **John Mattick**, University of Queensland, Brisbane, Australia
- **Fabrício Costa**, Northwestern University's Feinberg School of Medicine, Chicago, USA
- **Ing-Marie Olsson**, Umetrics AB, Malmö, Sweden